ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 787 – 814
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Adaptive estimation of the transition density of a particular hidden Markov chain

Claire Lacour

MAP5, Université Paris 5, 45 rue des Saints-Pères 75270 Paris Cedex 06, France

Received 29 June 2006
Available online 18 April 2007

Abstract

We study the following model of hidden Markov chain: $Y_i = X_i + \varepsilon_i, \; i = 1, \ldots, n+1$ with (X_i) a real-valued positive recurrent and stationary Markov chain, and $(\varepsilon_i)_{1 \leqslant i \leqslant n+1}$ a noise independent of the sequence (X_i) having a known distribution. We present an adaptive estimator of the transition density based on the quotient of a deconvolution estimator of the density of X_i and an estimator of the density of (X_i, X_{i+1}). These estimators are obtained by contrast minimization and model selection. We evaluate the L^2 risk and its rate of convergence for ordinary smooth and supersmooth noise with regard to ordinary smooth and supersmooth chains. Some examples are also detailed.

AMS 2000 subject classification: 62G05; 62H12; 62M05

Keywords: Hidden Markov chain; Transition density; Nonparametric estimation; Deconvolution; Model selection; Rate of convergence

1. Introduction

Let us consider the following model:

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \ldots, n+1, \tag{1}$$

where $(X_i)_{i \geqslant 1}$ is an irreducible and positive recurrent Markov chain and $(\varepsilon_i)_{i \geqslant 1}$ is a noise independent of $(X_i)_{i \geqslant 1}$. We assume that $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed random variables with known distribution. This model belongs to the class of hidden Markov models.

E-mail address: lacour@math-info.univ-paris5.fr.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2007.04.006

The hidden Markov models constitute a very famous class of discrete-time stochastic processes, with many applications in various areas such as biology, speech recognition or finance. For a general reference on these models, we refer to Cappé et al. [9]. Here, we study a simple model of HMM where the noise is additive (which obviously allows to deal also with multiplicative noise by use of logarithm). In standard HMM, it is assumed that the joint density of (X_i, Y_i) has a parametric form and the aim is then to infer the parameter from the observations $Y_1, \dots, Y_n$, generally by maximizing the likelihood. For this type of study, we can cite among others Baum and Petrie [3], Leroux [25], Bakry et al. [1], Bickel et al. [4], Jensen and Petersen [20], Douc et al. [16].

Here, we are interested in a nonparametric approach of the estimation of the hidden chain transition. A nonparametric model is particularly useful in the financial field (for instance in stochastic volatility model) where the form of the chain, which is usually derived from a diffusion, can be entirely unknown. So we assume that the Markov chain law is entirely unknown. Matias [27] and Butucea and Matias [7] considered the semiparametric problem where X_i follows an unknown distribution and the emission distribution has an unknown variance. The identifiability requires then for the signal density to be less regular than the density of the noise. Here, we assume that all regularities (in the sense defined below) for both distributions are possible but the noise distribution is completely known.

Our model is then a convolution model but with dependent variables X_i. The estimation of the density of X_i from the observations $Y_1, \dots, Y_n$ when the X_i's are i.i.d. (the so-called convolution model) has been extensively studied, see e.g. Caroll and Hall [10], Fan [18], Stefanski [29], Pensky and Vidakovic [28], Comte et al. [14]. However, very few authors study the case where (X_i) is a Markov chain. We can cite Dorea and Zhao [15] who estimate the density of Y_i in a very general context of HMM, Masry [26] who is interested in the estimation of the multivariate density in a mixing framework and Clémençon [12] who estimates the stationary density and the transition density of the hidden chain in the model (1). More precisely, he introduces an estimator of the transition density based on the thresholding of a wavelet-vaguelette decomposition and he studies its performance in the case of an ordinary smooth noise, that is with a polynomial decay of its Fourier transform.

Here, we are also interested in the estimation of the transition density of (X_i) but we consider a larger class of noise distributions. In Clémençon [12] there is no study of supersmooth noise (i.e. with exponentially decreasing Fourier transform), as with the Gaussian distribution. However, the study of such noise is essential for the applications and give interesting rates of convergence, in particular when the chain density is also supersmooth. In the present paper, the four cases (ordinary smooth or supersmooth noise with ordinary smooth or supersmooth chain) are considered.

The aim of this paper is to estimate the transition density Π of the Markov chain (X_i) from the observations $Y_1, \dots, Y_n$. To do this, we assume that the regime is stationary and we note that $\Pi = F/f$ where F is the density of (X_i, X_{i+1}) and f the stationary density. The estimation of f comes down to a problem of deconvolution, as does the estimation of F. We use contrast minimization and a model selection method inspired by Barron et al. [2] to find adaptive estimators of f and F. Our estimator of Π is then the quotient of the two previous estimators. Note that it is worth finding an adaptive estimator, i.e. an estimator whose risk automatically achieves the minimax rates, because the regularity of the densities f and F is generally very hard to compute, even if the chain can be fully described (as it is the case for a diffusion or an autoregressive process).

We study the performance of our estimator by computing the rate of convergence of the integrated risk. We improve the result of Clémençon [12] (case of an ordinary smooth noise) since we obtain the minimax rate without logarithmic loss. Moreover, we observe noticeable rates of convergence when both the noise and the chain are supersmooth.

The paper is organized as follows. Section 2 is devoted to notations and assumptions while the estimation procedure is developed in Section 3. After describing the projection spaces to which the estimators belong, we define separately the estimator of the stationary density f, the one of the joint density F and in the end the estimator $\tilde{\Pi}$ of the transition density. Section 4 states the results obtained for our estimators. To illustrate the theorems, some examples are provided in Section 5 as the AR(1) model, the Cox–Ingersoll–Ross process or the stochastic volatility model. The proofs are to be found in Section 6.

2. Notations and assumptions

For the sake of clarity, we use lowercase letters for dimension 1 and capital letters for dimension 2. For a function $t : \mathbb{R} \mapsto \mathbb{R}$, we denote by $\|t\|$ the L^2 norm that is $\|t\|^2 = \int_{\mathbb{R}} t^2(x)\,dx$. The Fourier transform t^* of t is defined by

$$t^*(u) = \int e^{-ixu} t(x)\,dx.$$

Note that the function t is the inverse Fourier transform of t^* and can be written $t(x) = 1/(2\pi)\int e^{ixu} t^*(u)\,du$. Finally, the convolution product is defined by $(t*s)(x) = \int t(x-y)s(y)\,dy$.

In the same way, for a function $T : \mathbb{R}^2 \mapsto \mathbb{R}$, $\|T\|^2 = \iint_{\mathbb{R}^2} T^2(x,y)\,dx\,dy$ and

$$T^*(u,v) = \int\int e^{-ixu-iyv} T(x,y)\,dx\,dy,$$
$$(T*S)(x,y) = \int\int T(x-z, y-w)S(z,w)\,dz\,dw.$$

We denote by $t \otimes s$ the function: $(x,y) \mapsto (t \otimes s)(x,y) = t(x)s(y)$.

The density of ε_i is named q and is known. We denote by p the unknown density of Y_i. We have $p = f * q$ and then $p^* = f^* q^*$. Similarly, if P is the density of (Y_i, Y_{i+1}), then $P = F * (q \otimes q)$ and $P^*(u,v) = F^*(u,v)q^*(u)q^*(v)$.

Now the assumptions on the model are the following:

A1. Function q^* never vanishes.

A2. There exist $s \geqslant 0$, $b > 0$, $\gamma \in \mathbb{R}$ ($\gamma > 0$ if $s = 0$) and $k_0, k_1 > 0$ such that

$$k_0(x^2+1)^{-\gamma/2}\exp(-b|x|^s) \leqslant |q^*(x)| \leqslant k_1(x^2+1)^{-\gamma/2}\exp(-b|x|^s).$$

A3. The chain is stationary with (unknown) density f.

A4. The chain is geometrically β-mixing ($\beta_q \leqslant Me^{-\theta q}$), or arithmetically β-mixing ($\beta_q \leqslant Mq^{-\theta}$) with $\theta > 8$.

That condition is verified as soon as the chain is uniformly ergodic. A definition of the β-mixing coefficients (in general and in the case of a Markov chain) can be found in Doukhan [17]. A lot of Markov chains satisfy Assumptions A4, see examples in Section 2.2 in Lacour [23].

In the sequel we consider the following smoothness spaces:

$$\mathcal{A}_{\delta,r,a}(l) = \left\{ f \text{ density on } \mathbb{R} \text{ and } \int |f^*(x)|^2 (x^2+1)^{\delta} \exp(2a|x|^r)\,dx \leqslant l \right\},$$

with $r \geqslant 0$, $a > 0$, $\delta \in \mathbb{R}$ ($\delta > 1/2$ if $r = 0$), $l > 0$ and

$$\mathbb{A}_{\Delta,R,A}(L) = \left\{ F \text{ density on } \mathbb{R}^2 \text{ and } \iint |F^*(x,y)|^2 (x^2+1)^{\Delta}(y^2+1)^{\Delta} \exp(2A(|x|^R + |y|^R))\,dx\,dy \leqslant L \right\},$$

with $R \geqslant 0$, $A > 0$, $\Delta \in \mathbb{R}$ ($\Delta > 1/2$ if $R = 0$), $L > 0$.

When $r > 0$ (respectively, $R > 0$) the function f (resp. F) is known as supersmooth, and as ordinary smooth otherwise. In the same way, the noise distribution is called ordinary smooth if $s = 0$ and supersmooth otherwise. The spaces of ordinary smooth functions correspond to classic Sobolev classes, while supersmooth functions are infinitely differentiable. It includes for example normal ($r = 2$) and Cauchy ($r = 1$) densities.

It is worth noting that as F is the density of (X_i, X_{i+1}), the two directions play a similar role. Thus, there is no use considering more general functional spaces for F, like anisotropic ones (see [24]).

3. Estimation procedure

Since $\Pi = F/f$ we proceed in three steps to estimate the transition density Π. First we find an estimator $\tilde{f}$ of f (see Section 3.2). Then we estimate F by $\tilde{F}$ (see Section 3.3). And finally we estimate Π with the quotient $\tilde{F}/\tilde{f}$ (Section 3.4).

All estimators defined here are projection estimators. We therefore start with describing the projection spaces.

3.1. Projection spaces

Let us consider the function

$$\varphi(x) = \sin(\pi x)/(\pi x)$$

and, for m in $\mathbb{N}^*$, j in $\mathbb{Z}$, $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j)$. Note that $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ is an orthonormal basis of the space of integrable functions having a Fourier transform with compact support included into $[-\pi m, \pi m]$. In the sequel, we use the following notations:

$$S_m = \mathrm{Span}\{\varphi_{m,j}\}_{j\in\mathbb{Z}}, \quad \mathbb{S}_m = \mathrm{Span}\{\varphi_{m,j} \otimes \varphi_{m,k}\}_{j,k\in\mathbb{Z}}.$$

These spaces have particular properties, which are a consequence of the first point of Lemma 3 (see Section 6.8):

$$\forall t \in S_m \;\; \|t\|_\infty \leqslant \sqrt{m}\|t\|, \quad \forall T \in \mathbb{S}_m \;\; \|T\|_\infty \leqslant m\|T\|, \tag{2}$$

where $\|t\|_\infty = \sup_{x\in\mathbb{R}} |t(x)|$ and $\|T\|_\infty = \sup_{(x,y)\in\mathbb{R}^2} |T(x,y)|$.

3.2. Estimation of f

Here, we estimate f, which is the density of the X_i's. It is the classic deconvolution problem. We choose to estimate f by minimizing a contrast. The standard contrast in density estimation is $(1/n)\sum_{i=1}^n [\|t\|^2 - 2t(X_i)]$. It is not possible to use this contrast here since we do not observe $X_1, \ldots, X_n$. Only the noisy data $Y_1, \ldots, Y_n$ are available. That is why we use the following lemma.

Lemma 1. *For all function t, let v_t be the inverse Fourier transform of $t^*/q^*(-.)$, i.e.*

$$v_t(x) = \frac{1}{2\pi}\int e^{ixu}\frac{t^*(u)}{q^*(-u)}\,du.$$

Then, for all $1 \leqslant k \leqslant n$,

(1) $\mathbb{E}[v_t(Y_k)|X_1,\ldots,X_n] = t(X_k)$,
(2) $\mathbb{E}[v_t(Y_k)] = \mathbb{E}[t(X_k)]$.

The second assertion in Lemma 1 is an obvious consequence of the first one and leads us to consider the following contrast:

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}[\|t\|^2 - 2v_t(Y_i)] \quad \text{with } v_t^*(u) = \frac{t^*(u)}{q^*(-u)}. \tag{3}$$

Indeed, since $t(X_i)$ and $v_t(Y_i)$ have the same expectation, it is natural to replace the unknown quantity $t(X_i)$ in the contrast by $v_t(Y_i)$.

We can observe that $\mathbb{E}\gamma_n(t)=(1/n)\sum_{i=1}^{n}[\|t\|^2-2\mathbb{E}[v_t(Y_i)]]=(1/n)\sum_{i=1}^{n}[\|t\|^2-2\mathbb{E}[t(X_i)]]= \|t\|^2 - 2\int tf = \|t-f\|^2 - \|f\|^2$ and then minimizing $\gamma_n(t)$ comes down to minimizing the distance between t and f. So we define

$$\hat{f}_m = \arg\min_{t\in S_m}\gamma_n(t) \tag{4}$$

or, equivalently,

$$\hat{f}_m = \sum_{j\in\mathbb{Z}}\hat{a}_j\varphi_{m,j} \quad \text{with } \hat{a}_j = \frac{1}{n}\sum_{i=1}^{n}v_{\varphi_{m,j}}(Y_i).$$

It is sufficient to differentiate the contrast to obtain this expression of the estimator. Actually, we should define $\hat{f}_m = \sum_{|j|\leqslant K_n}\hat{a}_j\varphi_{m,j}$ because we can estimate only a finite number of coefficients. If K_n is suitably chosen, it does not change the rate of convergence since the additional terms can be made negligible. For the sake of simplicity, we let the sum over $\mathbb{Z}$. For an example of detailed truncation see Comte [14].

Let f_m be the orthogonal projection of f on S_m, then

$$f_m = \sum_{j\in\mathbb{Z}}\left(\int f\varphi_{m,j}\right)\varphi_{m,j} = \sum_{j\in\mathbb{Z}}\mathbb{E}(\hat{a}_j)\varphi_{m,j}.$$

Conditionally to (X_i), the variance or stochastic error is

$$\begin{aligned}\mathbb{E}[\|\hat{f}_m - f_m\|^2|X_1,\ldots,X_n] &= \mathbb{E}\left[\sum_j(\hat{a}_j - \mathbb{E}(\hat{a}_j))^2|X_1,\ldots,X_n\right]\\ &= \sum_j \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}v_{\varphi_{m,j}}(Y_i)|X_1,\ldots,X_n\right] \leqslant \frac{\|\sum_j v_{\varphi_{m,j}}^2\|_\infty}{n}\end{aligned} \tag{5}$$

since $Y_1, \ldots, Y_n$ are independent conditionally to (X_i). Then, it follows from Lemma 3 (see Section 6.8) that $\| \sum_j v^2_{\varphi_{m,j}} \|_\infty = \Delta(m)$ where

$$\Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |q^*(u)|^{-2}\, du, \tag{6}$$

with q^* the characteristic function of the noise (ε_i). This implies that the order of the variance is $\Delta(m)/n$. That is why we introduce

$$\mathcal{M}_n = \left\{ m \geqslant 1,\ \frac{\Delta(m)}{n} \leqslant 1 \right\}.$$

To complete the estimation, we choose the best estimator among the collection $(\hat{f}_m)_{m\in\mathcal{M}_n}$. To do this we select the model which minimizes the following penalized criterion. Let

$$\hat{m} = \arg \min_{m\in\mathcal{M}_n} \{\gamma_n(\hat{f}_m) + \mathrm{pen}(m)\}$$

where pen is a penalty term to be specified later (see Theorem 1). Finally, we define $\tilde{f} = \hat{f}_{\hat{m}}$ our estimator of the stationary density.

3.3. Estimation of the density F of (X_i, X_{i+1})

We proceed similarly to the estimation of *f*. To define the contrast to minimize, we use the following lemma:

Lemma 2. *For all function T, let V_T be the inverse Fourier transform of $T^*/(q^* \otimes q^*)(-.)$, i.e.*

$$V_T(x, y) = \frac{1}{4\pi^2} \iint e^{ixu+iyv} \frac{T^*(u, v)}{q^*(-u)q^*(-v)}\, du\, dv.$$

Then, for all $1 \leqslant k \leqslant n$,

(1) $\mathbb{E}[V_T(Y_k, Y_{k+1})|X_1, \ldots, X_{n+1}] = T(X_k, X_{k+1})$,
(2) $\mathbb{E}[V_T(Y_k, Y_{k+1})] = \mathbb{E}[T(X_k, X_{k+1})]$.

We can now adapt contrast (3) to the bivariate case. For any function T in $L^2(\mathbb{R}^2)$, we define the contrast

$$\Gamma_n(T) = \frac{1}{n} \sum_{i=1}^{n} [\|T\|^2 - 2V_T(X_i, X_{i+1})]$$

whose expectation is equal to $\|T\|^2 - 2/n \sum_{k=1}^{n} \mathbb{E}[T(X_k, X_{k+1})] = \|T - F\|^2 - \|F\|^2$. As previously, we can define an estimator by minimizing the contrast function.

$$\hat{F}_m = \arg \min_{T\in\mathbb{S}_m} \Gamma_n(T). \tag{7}$$

By differentiating Γ_n, we obtain

$$\hat{F}_m(x, y) = \sum_{j,k} \hat{A}_{j,k}\varphi_{m,j}(x)\varphi_{m,k}(y) \quad \text{with } \hat{A}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} V_{\varphi_{m,j}\otimes\varphi_{m,k}}(Y_i, Y_{i+1}).$$

We choose again not to truncate the estimator for the sake of simplicity.

We have defined a collection of estimators $\{\hat{F}_m\}_{m\in\mathbb{M}_n}$ where we set

$$\mathbb{M}_n = \left\{ m \geqslant 1, \ \frac{\Delta^2(m)}{n} \leqslant 1 \right\},$$

with $\Delta(m)$ defined by (6). Indeed, as $V_{t\otimes s}(x, y) = v_t(x)v_s(y)$, the variance of the estimator $\hat{F}_m$ is now of order $\Delta^2(m)/n$ (see (5)). To define an adaptive estimator we have to select the best model m. So let

$$\hat{M} = \arg\min_{m\in\mathbb{M}_n} \{\Gamma_n(\hat{F}_m) + \mathrm{Pen}(m)\},$$

where Pen is a penalty function which is specified in Theorem 2. Finally, we consider the estimator $\tilde{F} = \hat{F}_{\hat{M}}$.

3.4. *Estimation of* Π

Whereas the estimation of f and F is valid on the whole real line $\mathbb{R}$ or $\mathbb{R}^2$, we estimate Π on a compact set B^2 only, because we need a lower bound on the stationary density. More precisely, we need to set some additional assumptions:

A5. There exists a positive real f_0 such that $\forall x \in B, \ f(x) \geqslant f_0$.
A6. $\forall x \in B, \forall y \in B, \Pi(x, y) \leqslant \|\Pi\|_{B,\infty} < \infty$.

Now, since $\Pi(x, y) = F(x, y)/f(x)$ we set

$$\tilde{\Pi}(x, y) = \begin{cases} \dfrac{\tilde{F}(x, y)}{\tilde{f}(x)} & \text{if} |\tilde{F}(x, y)| \leqslant n|\tilde{f}(x)|, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Here, the truncation allows to avoid the too small values of $\tilde{f}$ in the quotient. Now we evaluate upper bounds for the risk of our estimators.

4. Results

Our first theorem regards the problem of deconvolution. This result may be put together with results of Comte et al. [14] in the i.i.d. case and of Comte et al. [13] in various mixing frameworks.

Theorem 1. *Under Assumptions* A1–A4, *consider the estimator* $\tilde{f} = \hat{f}_{\hat{m}}$ *where for each* m, $\hat{f}_m$ *is defined by* (4) *and* $\hat{m} = \arg\min_{m\in\mathcal{M}_n}\{\gamma_n(\hat{f}_m) + \mathrm{pen}(m)\}$ *with*

$$\mathrm{pen}(m) = k\frac{(\pi m)^{[s-(1-s)_+/2]_+}\Delta(m)}{n},$$

where k is a constant depending only on k_0, k_1, b, γ, s. Then there exists $C > 0$ such that

$$\mathbb{E}\|\tilde{f} - f\|^2 \leqslant 4 \inf_{m\in\mathcal{M}_n} \{\|f_m - f\|^2 + \mathrm{pen}(m)\} + \frac{C}{n},$$

where f_m is the orthogonal projection of f on S_m.

The penalty is close to the variance order. It implies that the obtained rates of convergence are minimax in most cases. More precisely, the rates are given in the following corollary where $\lceil x \rceil$ denotes the ceiling function, i.e. the smallest integer larger than or equal to x.

Corollary 1. *Under Assumptions of Theorem* 1, *if f belongs to* $\mathcal{A}_{\delta,r,a}(l)$, *then*

- *If* $r = 0$ *and* $s = 0$, $\mathbb{E}\|\tilde{f} - f\|^2 \leqslant Cn^{-\frac{2\delta}{2\delta+2\gamma+1}}$.
- *If* $r = 0$ *and* $s > 0$, $\mathbb{E}\|\tilde{f} - f\|^2 \leqslant C(\ln n)^{-2\delta/s}$.
- *If* $r > 0$ *and* $s = 0$, $\mathbb{E}\|\tilde{f} - f\|^2 \leqslant C\frac{(\ln n)^{(2\gamma+1)/r}}{n}$.
- *If* $r > 0$ *and* $s > 0$
  - *if* $r < s$ *and* $k = \lceil (s/r-1)^{-1} \rceil - 1$, *there exist reals* b_i *such that*

$$\mathbb{E}\|\tilde{f} - f\|^2 \leqslant C(\ln n)^{-2\delta/s} \exp\left[\sum_{i=0}^{k} b_i (\ln n)^{(i+1)r/s-i}\right]$$

  - *if* $r = s$, *if* $\xi = [2\delta b + (s - 2\gamma - 1 - [s - (1-s)_+/2]_+)a]/[(a+b)s]$

$$\mathbb{E}\|\tilde{f} - f\|^2 \leqslant Cn^{-a/(a+b)}(\ln n)^{-\xi}$$

  - *if* $r > s$ *and* $k = \lceil (r/s-1)^{-1} \rceil - 1$, *there exist reals* d_i *such that*

$$\mathbb{E}\|\tilde{f} - f\|^2 \leqslant C\frac{(\ln n)^{(1+2\gamma-s+[s-(1-s)_+/2]_+)/r}}{n} \exp\left[-\sum_{i=0}^{k} d_i (\ln n)^{(i+1)s/r-i}\right].$$

These rates are the same as those obtained in the case of i.i.d. variables X_i; they are studied in detail in Comte et al. [14]. In this case, the rates $n^{-\frac{2\delta}{2\delta+2\gamma+1}}$ ($r = s = 0$), $(\ln n)^{-2\delta/s}$ ($r = 0$, $s > 0$) and $(\ln n)^{(2\gamma+1)/r}/n$ ($s = 0, r > 0$) are proved to be optimal by Fan [18] (first two cases) and Butucea [6] (third case) for i.i.d. variables. If $r > 0$ and $s > 0$, we find the original rates obtained in Lacour [22], proved as being optimal for $0 < r < s$ in Butucea and Tsybakov [8]. In the other cases, we can compare the results of Theorem 1 to the one obtained with a nonadaptive estimator. There is a loss only in the case $r \geqslant s > 1/3$ where a logarithmic term is added. But in this case, the rates are faster than any power of logarithm.

Now let us study the risk for our estimator of the joined density F.

Theorem 2. *Under Assumptions* A1–A4, *consider the estimator* $\tilde{F} = \hat{F}_{\hat{M}}$ *where for each* m, $\hat{F}_m$ *is defined by* (7) *and* $\hat{M} = \arg\min_{m \in \mathbb{M}_n} \{\Gamma_n(\hat{F}_m) + \text{Pen}(m)\}$ *with*

$$\text{Pen}(m) = K\frac{(\pi m)^{[s-(1-s)_+]_+}\Delta^2(m)}{n},$$

where K *is a constant depending only on* k_0, k_1, b, γ, s. *Then there exists* $C > 0$ *such that*

$$\mathbb{E}\|\tilde{F} - F\|^2 \leqslant 4 \inf_{m \in \mathbb{M}_n} \{\|F_m - F\|^2 + \text{Pen}(m)\} + \frac{C}{n},$$

where F_m *is the orthogonal projection of* F *on* $\mathbb{S}_m$.

The bases derived from the sine cardinal function are adapted to the estimation on the whole real line. The proof of Theorem 2 actually contains the proof of another result (see Proposition 2 in Section 6): the estimation of a bivariate density in a mixing framework on $\mathbb{R}^2$ and not only on a compact set. In this case of the absence of noise ($\varepsilon = 0$), we obtain the same result with the penalty $\mathrm{Pen}(m) = K_0(\sum_k \beta_{2k})m^2/n$. This limit case gives the mixing coefficients back in the penalty, as it always appears in this kind of estimation (see e.g. [30]).

It is then significant that in the presence of noise the penalty contains neither any mixing term nor any unknown quantity. It is entirely computable since it depends only on the characteristic function q^* of the noise which is known.

Theorem 2 enables us to give rates of convergence for the estimation of F.

Corollary 2. *Under Assumptions of Theorem 2, if F belongs to $\mathbb{A}_{\Delta,R,A}(L)$, then*

- *If $R = 0$ and $s = 0$,* $\mathbb{E}\|\tilde{F} - F\|^2 \leqslant C n^{-\frac{2\Delta}{2\Delta+4\gamma+2}}$.
- *If $R = 0$ and $s > 0$,* $\mathbb{E}\|\tilde{F} - F\|^2 \leqslant C(\ln n)^{-2\Delta/s}$.
- *If $R > 0$ and $s = 0$,* $\mathbb{E}\|\tilde{F} - F\|^2 \leqslant C\frac{(\ln n)^{(4\gamma+2)/R}}{n}$.
- *If $R > 0$ and $s > 0$*
    - *if $R < s$ and $k = \lceil (s/R - 1)^{-1} \rceil - 1$, there exist reals b_i such that*

$$\mathbb{E}\|\tilde{F} - F\|^2 \leqslant C(\ln n)^{-2\Delta/s} \exp\left[\sum_{i=0}^{k} b_i (\ln n)^{(i+1)R/s-i}\right]$$

    - *if $R = s$ if $\xi = [4\Delta b + (2s - 4\gamma - 2 - [s - (1 - s)_+]_+)A]/[(A + 2b)s]$*

$$\mathbb{E}\|\tilde{F} - F\|^2 \leqslant C n^{-A/(A+2b)} (\ln n)^{-\xi}$$

    - *if $R > s$ and $k = \lceil (R/s - 1)^{-1} \rceil - 1$, there exist reals d_i such that*

$$\mathbb{E}\|\tilde{F} - F\|^2 \leqslant C\frac{(\ln n)^{(2+4\gamma-2s+[s-(1-s)_+]_+)/R}}{n} \exp\left[-\sum_{i=0}^{k} d_i (\ln n)^{(i+1)s/R-i}\right].$$

The rates of convergence look like the one of Corollary 1 with modifications due to the bivariate nature of F. We can compare this result to the one of Clémençon [12] who studies only the case where $R = 0$ and $s = 0$. He shows that the minimax lower bound in that case is $n^{-\frac{2\Delta}{2\Delta+4\gamma+2}}$, so our procedure is optimal, whereas his estimator has a logarithmic loss for the upper bound. We remark that if $s > 0$ (supersmooth noise), the rate is logarithmic for F belonging to a classic ordinary smooth space. But if F is also supersmooth, better rates are recovered.

Except in the case where $R = 0$ and $s = 0$, there is, to our knowledge, no lower bound available for this estimation. We can, however, evaluate the performance of this estimator by comparing it with a nonadaptive estimator. If the smoothness of F is known, a value of m depending on R and Δ which minimizes the risk $\|F - F_m\|^2 + \Delta(m)^2/n$ can be exhibited and then some rates of convergence for this nonadaptive estimator are obtained. As soon as $s \leqslant 1/2$ (i.e. $[s-(1-s)_+]_+ = 0$), the penalty is $\Delta(m)^2/n$ and then the adaptive estimator recovers the same rates of convergence as those of a nonadaptive estimator if the regularity of F were known. It automatically minimizes the risk without prior knowledge on the regularity of F and there is no loss in the rates. If $s > 1/2$ a loss can appear but is not systematic. If $R < s$, the rate of convergence is unchanged since the bias dominates. It is only when $R \geqslant s > 1/2$ that an additional logarithmic term appears. But in this case the risk decreases faster than any logarithmic function so that the loss is negligible.

We can now state the main result regarding the estimation of the transition density Π.

Theorem 3. *Under Assumptions* A1–A6, *consider the estimator* $\tilde{\Pi}$ *defined in* (8). *We assume that* f *belongs to* $\mathcal{A}_{\delta,r,a}(l)$ *and that we browse only the models* $m \in \mathcal{M}_n$ *such that*

$$m \geqslant \ln\ln n \quad \text{and} \quad m\Delta(m) \leqslant \frac{n}{(\ln n)^2} \tag{9}$$

to define $\tilde{f}$. *Then* $\tilde{\Pi}$ *verifies, for n large enough,*

$$\mathbb{E}\|\tilde{\Pi} - \Pi\|_B^2 \leqslant C_1 \mathbb{E}\|\tilde{F} - F\|^2 + C_2 \mathbb{E}\|\tilde{f} - f\|^2 + \frac{C}{n},$$

where $\|T\|_B^2 = \iint_{B^2} T^2(x, y)\, dx\, dy$.

Note that, unlike in Theorems 1 and 2, this result is asymptotic. It states that the rate of convergence for Π is no larger than the maximum of the rates of f and F. The restrictions (9) do not modify the conclusion of Theorem 1 and the resulting rates of convergence. Thus if f and F have the same regularity, the rates of convergence for Π are those of F, given in Corollary 2.

If $s = 0$ i.e. if ε_i is ordinary smooth, then the rates of convergence are polynomial; moreover, they are near the parametric rate $1/n$ if R and r are positive. In the other hand the smoother the error distribution, the harder the estimation. In the case of a supersmooth noise, the rates are logarithmic if f or F is ordinary smooth but faster than any power of logarithm if the hidden chain has supersmooth densities. The exact rates depend on all regularities γ, s, δ, r, Δ, R and are very tedious to write. That is why we prefer to give some detailed examples.

5. Examples

In this section, we give some examples to illustrate the previous results. In nonparametric examples, the quantities that allow to compute the rates of convergence, i.e. the regularities of the densities, remain unknown. It is besides an advantage of the procedure, not to need such information to reach good rates.

So the following models are parametric, but it is well known that in the case where the state spaces of the hidden chains are not finite, nor bounded, classical parametric estimation is not proved to perform well.

5.1. Autoregressive process of order 1

Let us study the case where the Markov chain is defined by

$$X_{n+1} = \alpha X_n + \beta + \eta_{n+1},$$

where the η_n's are i.i.d. centered Gaussian with variance σ^2. This chain is irreducible, Harris recurrent and geometrically β-mixing. The stationary distribution is Gaussian with mean $\beta/(1-\alpha)$ and variance $\sigma^2/(1-\alpha^2)$. So

$$f^*(u) = \exp\left[-iu\left(\frac{\beta}{1-\alpha}\right) - \frac{\sigma^2}{2(1-\alpha^2)}u^2\right]$$

and then bias computing gives $\delta = 1/2, r = 2$. The function F is the density of a Gaussian vector with mean $(\beta/(1-\alpha), \beta/(1-\alpha))$ and variance matrix $\sigma^2/(1-\alpha^2)\begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$. So

$$F^*(u,v) = \exp\left[-i(u+v)\left(\frac{\beta}{1-\alpha}\right) - \frac{\sigma^2}{2(1-\alpha^2)}(u^2+v^2+2\alpha uv)\right]$$

and $\Delta = 1/2$, $R = 2$.

We can compute the rates of convergence for different kinds of noise ε. If ε has a Laplace distribution, $q^*(u) = 1/(1+u^2)$ so $s = 0$, $\gamma = 2$. In this case, Corollary 1 gives $\mathbb{E}\|\tilde{f} - f\|^2 \leqslant C(\ln n)^{5/2}/n$ and $\mathbb{E}\|\tilde{F} - F\|^2 \leqslant C(\ln n)^5/n$. Consequently,

$$\mathbb{E}\|\tilde{\Pi} - \Pi\|_B^2 \leqslant C\frac{(\ln n)^5}{n},$$

with B an interval $[-d, d]$. This rate is close to the parametric rate $1/n$; it is due to the great smoothness of the chain compared with that of error.

If now ε has a normal distribution with variance τ^2, then we compute

$$\mathbb{E}\|\tilde{\Pi} - \Pi\|_B^2 \leqslant Cn^{-\frac{\sigma^2}{\sigma^2+2\tau^2}}(\ln n)^{-\frac{\tau^2}{\sigma^2+2\tau^2}}.$$

5.2. *Cox–Ingersoll–Ross process*

Another example is given by $X_n = R_{n\tau}$ with τ a fixed sampling interval and R_t the so-called Cox–Ingersoll–Ross process defined by

$$dR_t = (2\theta R_t + \kappa\sigma_0^2)\,dt + 2\sigma_0\sqrt{R_t}\,dW_t, \quad \theta < 0, \quad \kappa \in \{2, 3, \ldots\}.$$

Following Chaleya-Maurel and Genon-Catalot [11], we observe that X_n is the square of the Euclidean norm of a κ-dimensional vector whose components are linear autoregressive processes of order 1. The stationary distribution is a Gamma distribution with parameter $\kappa/2$ and $|\theta|/\sigma_0^2$ so that

$$f^*(u) = \left(1 + iu\frac{\sigma_0^2}{|\theta|}\right)^{-\kappa/2}$$

and $r = 0$, $\delta = (\kappa-1)/2$. To compute the characteristic function of the joined density, we write

$$F^*(u,v) = \int \mathbb{E}[e^{-ivX_1}|X_0 = x]e^{-iux}f(x)\,dx.$$

Let $\beta^2 = \sigma_0^2(e^{2\theta\tau} - 1)/(2\theta)$. Then, conditionally to $X_0 = x$, $\beta^{-2}X_1$ is a noncentral chi-square $\chi'^2(e^{2\theta\tau}x/\beta^2, \kappa)$, so that

$$\mathbb{E}[e^{-ivX_1}|X_0 = x] = (1 + 2iv\beta^2)^{-\kappa/2}\exp\left(-\frac{ive^{2\theta\tau}x}{1+2iv\beta^2}\right).$$

This implies

$$F^*(u,v) = \left[1 - (1 - e^{2\theta\tau})\frac{\sigma_0^4}{\theta^2}uv + i\frac{\sigma_0^2}{|\theta|}(u+v)\right]^{-\kappa/2}$$

and $R = 0$, $\Delta = (\kappa - 1)/2$. Then, if for example the noise has a Gaussian distribution ($\gamma = 0$, $s = 2$), the rate of convergence is $(\ln n)^{(1-\kappa)/2}$. But this rate is faster if ε has a Gamma distribution with shape parameter α (so that $\gamma = \alpha$, $s = 0$): we obtain in this case $n^{(1-\kappa)/(\kappa+4\alpha+1)}$.

5.3. Stochastic volatility model

Our work allows to study some multiplicative models as the so-called stochastic volatility model in finance (see [19] for the links between the standard continuous-time SV models and the hidden Markov models). Let us consider

$$Z_n = U_n^{1/2} \eta_n,$$

where (U_n) is a nonnegative Markov chain, (η_n) a sequence of i.i.d. standard Gaussian variables, the two sequences being independent. Setting $X_n = \ln(U_n)$ and $\varepsilon_n = \ln(\eta_n^2)$ leads us back to our initial problem.

The noise distribution is the logarithm of a chi-square distribution and then verifies $q^*(x) = 2^{-ix}\Gamma(1/2 - ix)/\sqrt{\pi}$. Van Es et al. [31] show that $|q^*(x)| \sim_{+\infty} \sqrt{2}e^{-\pi|x|/2}$ and then s=1, γ=0.

In the general case, the logarithm of the hidden chain X_n derives from a regular sampling of a diffusion process with unknown drift and diffusion coefficients. Then the rate of convergence for the estimation of the transition depend on the smoothness of f and F. If $R = r = 0$, then $\mathbb{E}\|\tilde{\Pi} - \Pi\|_B^2 \leqslant C(\ln n)^{-2\delta}$. But if r and R are positive, better rates are recovered.

For example, we assume that the logarithm of the hidden chain X_n derives from a regular sampling of an Ornstein–Uhlenbeck process, i.e. $X_n = V_{n\tau}$ where V_t is defined by the equation

$$dV_t = \theta V_t\, dt + \sigma\, dB_t$$

with B_t a standard Brownian motion. Then all the assumptions are satisfied. Similarly to Section 5.1, the stationary distribution is Gaussian with mean 0 and variance $\sigma^2/2|\theta|$ and then $\delta = 1/2, r = 2$. In the same way F is the density of a centered Gaussian vector with variance matrix $\sigma^2/(2|\theta|)\begin{pmatrix} 1 & e^{\theta\tau} \\ e^{\theta\tau} & 1 \end{pmatrix}$ and then $\Delta = 1/2$, $R = 2$. We obtain the following rate of convergence on some interval $B = [-d, d]$:

$$\mathbb{E}\|\tilde{\Pi} - \Pi\|_B^2 \leqslant C\sqrt{\ln n}\,\frac{\exp[(\pi/\beta)\sqrt{\ln n}]}{n},$$

with $\beta^2 = \sigma^2(e^{2\theta\tau} - 1)/(2\theta)$.

6. Proofs

Here, we do not prove the results concerning the estimation of f. Indeed, they are similar to the ones concerning F (but actually simpler) and the ones of Comte et al. [14]. It is then sufficient to use corresponding proofs for F mutatis mutandis.

For the sake of simplicity, all constants in the following are denoted by C, even if they have different values.

6.1. Proof of Lemma 2

It is sufficient to prove the first assertion. First we write that $V_T(Y_k, Y_{k+1}) = 1/4\pi^2 \int e^{iY_k u + iY_{k+1}v} T^*(u,v)/q^*(-u)q^*(-v)\,du\,dv$ so that

$$\mathbb{E}[V_T(Y_k, Y_{k+1})|X_1, \ldots, X_{n+1}] = \frac{1}{4\pi^2}\int \mathbb{E}[e^{iY_k u + iY_{k+1}v}|X_1, \ldots, X_{n+1}]\frac{T^*(u,v)}{q^*(-u)q^*(-v)}\,du\,dv.$$

By using the independence between (X_i) and (ε_i), we compute

$$\begin{aligned}
&\mathbb{E}[e^{iY_k u + iY_{k+1}v}|X_1, \ldots, X_{n+1}] = \mathbb{E}[e^{iX_k u + iX_{k+1}v}e^{i\varepsilon_k u + i\varepsilon_{k+1}v}|X_1, \ldots, X_{n+1}]\\
&\quad = e^{iX_k u + iX_{k+1}v}\mathbb{E}[e^{i\varepsilon_k u}]\mathbb{E}[e^{i\varepsilon_{k+1}v}] = e^{iX_k u + iX_{k+1}v}\int e^{ixu}q(x)\,dx \int e^{iyv}q(y)\,dy\\
&\quad = e^{iX_k u + iX_{k+1}v}q^*(-u)q^*(-v).
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}[V_T(Y_k, Y_{k+1})|X_1, \ldots, X_{n+1}] &= \frac{1}{4\pi^2}\int e^{iX_k u + iX_{k+1}v}q^*(-u)q^*(-v)\frac{T^*(u,v)}{q^*(-u)q^*(-v)}\,du\,dv\\
&= \frac{1}{4\pi^2}\int e^{iX_k u + iX_{k+1}v}T^*(u,v)\,du\,dv = T(X_k, X_{k+1}).
\end{aligned}$$

6.2. Sketch of proof of Theorem 2

Let $m \in \mathcal{M}_n$. The definitions of $\hat{F}_m$ and $\hat{m}$ lead to the inequality

$$\Gamma_n(\hat{F}_{\hat{M}}) + \mathrm{Pen}(\hat{M}) \leqslant \Gamma_n(F_m) + \mathrm{Pen}(m). \tag{10}$$

Let

$$Z_n(T) = (1/n)\sum_{i=1}^{n}\left\{V_T(Y_i, Y_{i+1}) - \int T(x,y)F(x,y)\,dx\,dy\right\}. \tag{11}$$

It is easy to see that

$$\Gamma_n(\hat{F}_{\hat{M}}) - \Gamma_n(F_m) = \|\hat{F}_{\hat{M}} - F\|^2 - \|F_m - F\|^2 - 2Z_n(\hat{F}_{\hat{M}} - F_m)$$

so that (10) becomes

$$\begin{aligned}
\|\hat{F}_{\hat{M}} - F\|^2 &\leqslant \|F_m - F\|^2 + 2Z_n(\hat{F}_{\hat{M}} - F_m) + \mathrm{Pen}(m) - \mathrm{Pen}(\hat{M})\\
&\leqslant \|F_m - F\|^2 + 2\|\hat{F}_{\hat{M}} - F_m\| \sup_{T \in B(m,\hat{M})} Z_n(T) + \mathrm{Pen}(m) - \mathrm{Pen}(\hat{M}),
\end{aligned}$$

where $B(m, m') = \{T \in \mathbb{S}_m + \mathbb{S}_{m'}, \|T\| = 1\}$, The main step of the proof is then to control the term $\sup_{T\in B(m,\hat{M})} Z_n(T)$.

To deal with the supremum of the empirical process $Z_n(T)$, we will use an inequality of Talagrand stated in Lemma 5 (Section 6.8). This inequality is very powerful but can be applied only to sum of independent random variables. That is why we split $Z_n(T)$ into two processes.

$$Z_n(T) = Z_{n,1}(T) + Z_{n,2}(T),$$

with

$$\begin{cases} Z_{n,1}(T) = \dfrac{1}{n}\sum_{i=1}^{n}\{V_T(Y_i, Y_{i+1}) - \mathbb{E}[V_T(Y_i, Y_{i+1})|X_1, \ldots, X_{n+1}]\}, \\ Z_{n,2}(T) = \dfrac{1}{n}\sum_{i=1}^{n}\left\{T(X_i, X_{i+1}) - \int T(x, y)F(x, y)\, dx\, dy\right\}. \end{cases} \tag{12}$$

For the first process $Z_{n,1}(T)$, we return to independent variables remarking that, conditionally to $X_1, \ldots, X_{n+1}$, the variables (Y_{2i-1}, Y_{2i}) are independent (see Proposition 1).

For the other processes, we use the mixing assumption A4 to build auxiliary variables X_i^* which are approximation of the X_i and which constitute independent clusters of variables (see Proposition 2).

6.3. Detailed proof of Theorem 2

First, we introduce some auxiliary variables whose existence is ensured by Assumption A4 of mixing. In the case of arithmetical mixing, since $\theta > 8$, there exists a real c such that $0 < c < 1/2$ and $c\theta > 4$. We set in this case $q_n = \frac{1}{2}\lfloor n^c \rfloor$. In the case of geometrical mixing, we set $q_n = \frac{1}{2}\lfloor c\ln(n)\rfloor$ where c is a real larger than $4/\theta$.

For the sake of simplicity, we suppose that $n = 4p_nq_n$, with p_n an integer. Let for $i = 1, \ldots, n/2$, $V_i = (X_{2i-1}, X_{2i})$ and for $l = 0, \ldots, p_n - 1$, $A_l = (V_{2lq_n+1}, \ldots, V_{(2l+1)q_n})$, $B_l = (V_{(2l+1)q_n+1}, \ldots, V_{(2l+2)q_n})$. As in Viennet [32], by using Berbee's coupling Lemma, we can build a sequence (A_l^*) such that

$$\begin{cases} A_l \text{ and } A_l^* \text{ have the same distribution,} \\ A_l^* \text{ and } A_{l'}^* \text{ are independent if } l \neq l', \\ P(A_l \neq A_l^*) \leqslant \beta_{2q_n}. \end{cases}$$

In the same way, we build (B_l^*) and we define for any $l \in \{0, \ldots, p_n - 1\}$, $A_l^* = (V_{2lq_n+1}^*, \ldots, V_{(2l+1)q_n}^*)$, $B_l^* = (V_{(2l+1)q_n+1}^*, \ldots, V_{(2l+2)q_n}^*)$ so that the sequence $(V_1^*, \ldots, V_{n/2}^*)$ and then the sequence $(X_1^*, \ldots, X_n^*)$ are well defined. We can now define

$$\Omega^* = \{\forall i, 1 \leqslant i \leqslant n,\ X_i = X_i^*\}.$$

Then we split the risk into two terms:

$$\mathbb{E}(\|\tilde{F} - F\|^2) = \mathbb{E}(\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^*}) + \mathbb{E}(\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^{*c}}).$$

To pursue the proof, we observe that for all T, T'

$$\Gamma_n(T) - \Gamma_n(T') = \|T - F\|^2 - \|T' - F\|^2 - 2Z_n(T - T'),$$

where $Z_n(T)$ is defined by (11). Let us fix $m \in \mathbb{M}_n$ and denote by F_m the orthogonal projection of F on $\mathbb{S}_m$. Since $\Gamma_n(\tilde{F}) + \mathrm{Pen}(\hat{M}) \leqslant \Gamma_n(F_m) + \mathrm{Pen}(m)$, we have

$$\begin{aligned} \|\tilde{F} - F\|^2 &\leqslant \|F_m - F\|^2 + 2Z_n(\tilde{F} - F_m) + \mathrm{Pen}(m) - \mathrm{Pen}(\hat{M}) \\ &\leqslant \|F_m - F\|^2 + 2\|\tilde{F} - F_m\| \sup_{T \in B(m,\hat{M})} Z_n(T) + \mathrm{Pen}(m) - \mathrm{Pen}(\hat{M}), \end{aligned}$$

where, for all m, m', $B(m, m') = \{T \in \mathbb{S}_m + \mathbb{S}_{m'}, \|T\| = 1\}$. Then, using inequality $2xy \leqslant x^2/4 + 4y^2$,

$$\|\tilde{F} - F\|^2 \leqslant \|F_m - F\|^2 + \frac{1}{4}\|\tilde{F} - F_m\|^2 + 4 \sup_{T \in B(m,\hat{M})} Z_n^2(T) + \mathrm{Pen}(m) - \mathrm{Pen}(\hat{M}). \tag{13}$$

Using Lemma 2, $Z_n(T)$ can be split into two terms:

$$Z_n(T) = Z_{n,1}(T) + Z_{n,2}(T),$$

with $Z_{n,1}(T)$ and $Z_{n,2}(T)$ defined by (12). Now let $P_1(.,.)$ be a function such that for all m, m',

$$16P_1(m, m') \leqslant \mathrm{Pen}(m) + \mathrm{Pen}(m'). \tag{14}$$

Then (13) becomes

$$\begin{aligned}\|\tilde{F} - F\|^2 \leqslant{}& \|F_m - F\|^2 + \frac{1}{2}(\|\tilde{F} - F\|^2 + \|F - F_m\|^2) + 2\,\mathrm{Pen}(m)\\ &+8\left[\sup_{T \in B(m,\hat{M})} Z_{n,1}^2(T) - P_1(m, \hat{M})\right] + 8\left[\sup_{T \in B(m,\hat{M})} Z_{n,2}^2(T) - P_1(m, \hat{M})\right]\end{aligned}$$

which gives, by introducing a function $P_2(.,.)$,

$$\begin{aligned}\frac{1}{2}\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^*} \leqslant{}& \frac{3}{2}\|F_m - F\|^2 + 2\,\mathrm{Pen}(m) + 8\sum_{m' \in \mathbb{M}_n}\left[\sup_{T \in B(m,m')} Z_{n,1}^2(T) - P_1(m, m')\right]\\ &+8\sum_{m' \in \mathbb{M}_n}\left[\sup_{T \in B(m,m')} Z_{n,2}^2(T) - P_2(m, m')\right]_+ \mathbb{1}_{\Omega^*}\\ &+8\sum_{m' \in \mathbb{M}_n}[P_2(m, m') - P_1(m, m')].\end{aligned}$$

We now use the following propositions:

Proposition 1. *Let* $P_1(m, m') = C(q)(\pi m'')^{[s-(1-s)_+]_+}\Delta^2(m'')/n$ *where* $\Delta(m)$ *is defined in* (6) *and* $m'' = \max(m, m')$ *and* $C(q)$ *is a constant. Then, under assumptions of Theorem* 2, *there exists a positive constant C such that*

$$\sum_{m' \in \mathbb{M}_n} \mathbb{E}\left(\left[\sup_{T \in B(m,m')} Z_{n,1}^2(T) - P_1(m, m')\right]_+\right) \leqslant \frac{C}{n}. \tag{15}$$

Proposition 2. *Let* $P_2(m, m') = 96(\sum_k \beta_{2k})m''/n$ *where* $m'' = \max(m, m')$. *Then, under assumptions of Theorem* 2, *there exists a positive constant C such that*

$$\sum_{m' \in \mathbb{M}_n} \mathbb{E}\left(\left[\sup_{T \in B(m,m')} Z_{n,2}^2(T) - P_2(m, m')\right]_+ \mathbb{1}_{\Omega^*}\right) \leqslant \frac{C}{n}. \tag{16}$$

The definitions of the functions $P_1(m, m')$ and $P_2(m, m')$ given in Propositions 1 and 2 imply that there exists m_0 such that $\forall m' > m_0$, $P_1(m, m') \geqslant P_2(m, m')$. (If $s = 0 = \gamma$ (case of a null noise), it would be wrong and the penalty would then be $P_2(m, m')$ instead of $P_1(m, m')$.) Then

$$\sum_{m' \in \mathbb{M}_n} [P_2(m, m') - P_1(m, m')] \leqslant \sum_{m' \leqslant m_0} P_2(m, m') \leqslant \frac{C(m_0)}{n}. \tag{17}$$

Since $m''\Delta^2(m'') \leqslant m\Delta^2(m) + m'\Delta^2(m')$, condition (14) is verified with

$$\mathrm{Pen}(m) = 16C(q)(\pi m)^{[s-(1-s)_+]_+} \frac{\Delta^2(m)}{n}.$$

And finally, combining (17) and Propositions 1 and 2,

$$\mathbb{E}(\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^*}) \leqslant 4(\|F_m - F\|^2 + \mathrm{Pen}(m)) + \frac{C}{n}.$$

For the term $\mathbb{E}(\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^{*c}})$, recall that

$$\hat{F}_m(x, y) = \sum_{j,k} \hat{A}_{j,k} \varphi_{m,j}(x)\varphi_{m,k}(y) \quad \text{with } \hat{A}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} V_{\varphi_{m,j}\otimes\varphi_{m,k}}(Y_i, Y_{i+1}).$$

Thus, for any m in $\mathbb{M}_n$,

$$\begin{aligned}
\|\hat{F}_m\|^2 &= \sum_{j,k}\left[\frac{1}{n}\sum_{i=1}^{n} V_{\varphi_{m,j}\otimes\varphi_{m,k}}(Y_i, Y_{i+1})\right]^2 \leqslant \frac{1}{n^2}\sum_{j,k} n \sum_{i=1}^{n} V^2_{\varphi_{m,j}\otimes\varphi_{m,k}}(Y_i, Y_{i+1}) \\
&\leqslant \left\|\sum_{j,k} V^2_{\varphi_{m,j}\otimes\varphi_{m,k}}\right\|_\infty \leqslant \left\|\sum_{j} v^2_{\varphi_{m,j}}\right\|^2_\infty \leqslant \Delta^2(m)
\end{aligned} \tag{18}$$

using Lemma 3 (see Section 6.8). Then $\|\hat{F}_{\hat{M}}\|^2 \leqslant \Delta^2(\hat{M}) \leqslant n$ since $\hat{M}$ belongs to $\mathbb{M}_n$. And

$$\mathbb{E}\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^{*c}} \leqslant \mathbb{E}(2(\|\tilde{F}\|^2 + \|F\|^2)\mathbb{1}^{*c}_{\Omega}) \leqslant 2(n + \|F\|^2)P(\Omega^{*c}).$$

Using Assumption A4 in the geometric case, $\beta_{2q_n} \leqslant Me^{-\theta c \ln(n)} \leqslant Mn^{-\theta c}$ and, in the other case, $\beta_{2q_n} \leqslant M(2q_n)^{-\theta} \leqslant Mn^{-\theta c}$. Then $P(\Omega^{*c}) \leqslant 2p_n\beta_{2q_n} \leqslant nMn^{-c\theta}$. Since $c\theta > 4$, $P(\Omega^{*c}) \leqslant Mn^{-3}$, which implies $E(\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^{*c}}) \leqslant C/n^2$.

Finally we obtain

$$\begin{aligned}
\mathbb{E}\|\tilde{F} - F\|^2 &\leqslant \mathbb{E}(\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^*}) + E(\|\tilde{F} - F\|^2 \mathbb{1}_{\Omega^{*c}}) \\
&\leqslant 4(\|F_m - F\|^2 + \mathrm{Pen}(m)) + \frac{C}{n}.
\end{aligned}$$

This inequality holds for each $m \in \mathbb{M}_n$, so the result is proved.

6.4. Proof of Proposition 1

We start by isolating odd terms from even terms to avoid overlaps:

$$Z_{n,1}(T) = \tfrac{1}{2} Z^o_{n,1}(T) + \tfrac{1}{2} Z^e_{n,1}(T),$$

with

$$\begin{cases} Z^o_{n,1}(T) = \dfrac{2}{n} \sum\limits_{i=1, i \text{ odd}}^{n} \{V_T(Y_i, Y_{i+1}) - \mathbb{E}_X[V_T(Y_i, Y_{i+1})]\}, \\ Z^e_{n,1}(T) = \dfrac{2}{n} \sum\limits_{i=1, \text{ even}}^{n} \{V_T(Y_i, Y_{i+1}) - \mathbb{E}_X[V_T(Y_i, Y_{i+1})]\} \end{cases}$$

denoting by $\mathbb{E}_X$ the expectation conditionally to $X_1, \ldots, X_{n+1}$. It is sufficient to deal with the first term, the second one being similar. For each i, let $U_i = (Y_{2i-1}, Y_{2i})$, then

$$Z^o_{n,1}(T) = \frac{1}{n/2} \sum_{i=1}^{n/2} \{V_T(U_i) - \mathbb{E}_X[V_T(U_i)]\}.$$

Let us remark that conditionally to $X_1, \ldots, X_{n+1}$, the U_i's are independent. Thus, we can use the Talagrand inequality recalled in Lemma 5 (see Section 6.8). Note that if T belongs to $\mathbb{S}_m + \mathbb{S}_{m'}$, then T can be written $T_1 + T_2$ where T_1^* has its support in $[-\pi m, \pi m]^2$ and T_2^* has its support in $[-\pi m', \pi m']^2$. Then T belongs to $\mathbb{S}_{m''}$ where m'' is defined by

$$m'' = \max(m, m'). \tag{19}$$

Now let us compute M_1, H and v of the Talagrand's inequality.

(1) If T belongs to $B(m, m')$,

$$V_T(x, y) = \sum_{j,k} a_{jk} V_{\varphi_{m'',j} \otimes \varphi_{m'',k}}(x, y) = \sum_{j,k} a_{jk} v_{\varphi_{m'',j}}(x) v_{\varphi_{m'',k}}(y).$$

Thus $|V_T(x, y)|^2 \leqslant \sum_{j,k} |v_{\varphi_{m'',j}}(x) v_{\varphi_{m'',k}}(y)|^2$. So

$$\sup_{T \in B(m,m')} \|V_T\|_\infty^2 \leqslant \left\| \sum_{j,k} |v_{\varphi_{m'',j}}(x) v_{\varphi_{m'',k}}(y)|^2 \right\|_\infty \leqslant \left\| \sum_j |v_{\varphi_{m'',j}}|^2 \right\|_\infty^2 .$$

By using Lemma 3, $M_1 = \Delta(m'')$.

(2) To compute H^2, we write

$$\begin{aligned} \mathbb{E}_X \left(\sup_{T \in B(m,m')} (Z^o_{n,1})^2(T) \right) &\leqslant \mathbb{E}_X \left(\sum_{j,k} Z^o_{n,1}(\varphi_{m'',j} \otimes \varphi_{m'',k})^2 \right) \\ &\leqslant \sum_{j,k} \mathrm{Var}_X \left[\frac{2}{n} \sum_{i=1, i \text{ odd}}^{n} v_{\varphi_{m'',j}}(Y_i) v_{\varphi_{m'',k}}(Y_{i+1}) \right] \\ &\leqslant \sum_{j,k} \frac{4}{n^2} \sum_{i=1, i \text{ odd}}^{n} \mathrm{Var}_X [v_{\varphi_{m'',j}}(Y_i) v_{\varphi_{m'',k}}(Y_{i+1})] \end{aligned}$$

since, conditionally to $X_1, \ldots, X_{n+1}$, the U_i's are independent. And then

$$\mathbb{E}_X\left(\sup_{T\in B(m,m')} Z_{n,1}^{o2}(T)\right) \leqslant \sum_{j,k} \frac{4}{n^2} \sum_{i=1,i \text{ odd}}^{n} \mathbb{E}_X[v_{\varphi_{m'',j}}^2(Y_i)v_{\varphi_{m'',k}}^2(Y_{i+1})]$$
$$\leqslant \frac{4}{n^2} \sum_{i=1,i \text{ odd}}^{n} \left\|\sum_j |v_{\varphi_{m'',j}}|^2\right\|_\infty \left\|\sum_k |v_{\varphi_{m'',k}}|^2\right\|_\infty \leqslant \frac{2\Delta(m'')^2}{n}.$$

So we set $H = \sqrt{2}\Delta(m'')/\sqrt{n}$.

(3) We still have to find v. On the one hand

$$\mathrm{Var}_X[V_T(Y_k, Y_{k+1})] \leqslant \mathbb{E}_X\left[\left(\sum_{j,k} a_{jk} v_{\varphi_{m'',j}}(Y_k)v_{\varphi_{m'',k}}(Y_{k+1})\right)^2\right]$$
$$\leqslant \sum_{j,k} a_{jk}^2 \left\|\sum_j |v_{\varphi_{m'',j}}|^2\right\|_\infty \left\|\sum_k |v_{\varphi_{m'',k}}|^2\right\|_\infty$$

and so $v \geqslant \Delta(m'')^2$. On the other hand

$$\begin{aligned}
&\mathrm{Var}_X[V_T(Y_k, Y_{k+1})] \\
&\quad\leqslant \sum_{j_1,k_1}\sum_{j_2,k_2} a_{j_1k_1}a_{j_2k_2}\mathbb{E}_X[v_{\varphi_{m'',j_1}}v_{\varphi_{m'',j_2}}(Y_k)v_{\varphi_{m'',k_1}}v_{\varphi_{m'',k_2}}(Y_{k+1})] \\
&\quad\leqslant \sum_{j,k} a_{jk}^2 \sqrt{\sum_{j_1,k_1}\sum_{j_2,k_2} \mathbb{E}_X^2[v_{\varphi_{m'',j_1}}v_{\varphi_{m'',j_2}}(Y_k)v_{\varphi_{m'',k_1}}v_{\varphi_{m'',k_2}}(Y_{k+1})]} \\
&\quad\leqslant \sum_{j,k} a_{jk}^2 \sqrt{\sum_{j_1,j_2} \mathbb{E}_X^2[v_{\varphi_{m'',j_1}}v_{\varphi_{m'',j_2}}(Y_k)]\sum_{k_1,k_2}\mathbb{E}_X^2[v_{\varphi_{m'',k_1}}v_{\varphi_{m'',k_2}}(Y_{k+1})]},
\end{aligned} \tag{20}$$

using conditional independence. Now we use Lemma 3 to compute

$$\begin{aligned}
&\mathbb{E}_X[v_{\varphi_{m'',j_1}}v_{\varphi_{m'',j_2}}(Y_k)] \\
&\quad= \int (v_{\varphi_{m'',j_1}}v_{\varphi_{m'',j_2}})(X_k+x)q(x)\,dx \\
&\quad= \frac{m''}{4\pi^2}\int\int_{-\pi}^{\pi} \frac{e^{-ij_1v}e^{i(x+X_k)vm''}}{q^*(-vm'')}\,dv \int_{-\pi}^{\pi} \frac{e^{-ij_2u}e^{i(x+X_k)um''}}{q^*(-um'')}\,du\,q(x)\,dx \\
&\quad= \frac{m''}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{e^{-ij_1v-ij_2u}e^{iX_k(u+v)m''}}{q^*(-vm'')q^*(-um'')}\int e^{ix(u+v)m''}q(x)\,dx\,du\,dv.
\end{aligned}$$

If we set $W(u,v) = m''e^{iX_k(u+v)m''}q^*(-(u+v)m'')/[q^*(-vm'')q^*(-um'')]$, then $\mathbb{E}_X[v_{\varphi_{m'',j_1}}v_{\varphi_{m'',j_2}}(Y_k)]$ is the Fourier coefficient with order (j_1, j_2) of W. Using Parseval's formula

$$\sum_{j_1,j_2} \mathbb{E}_X^2[v_{\varphi_{m'',j_1}}v_{\varphi_{m'',j_2}}(Y_k)] = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |W(u,v)|^2\,du\,dv$$
$$= \frac{m''^2}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\frac{q^*(-(u+v)m'')}{q^*(-vm'')q^*(-um'')}\right|^2 du\,dv.$$

Now we apply the Schwarz inequality:

$$\begin{aligned}
&\sum_{j_1,j_2} \mathbb{E}_X^2[v_{\varphi_{m'',j_1}}(Y_k)] \\
&\quad\leqslant \frac{m''^2}{4\pi^2}\sqrt{\iint \frac{|q^*(-(u+v)m'')|^2}{|q^*(-um'')|^4}\,du\,dv}\sqrt{\iint \frac{|q^*(-(u+v)m'')|^2}{|q^*(-vm'')|^4}\,du\,dv} \\
&\quad\leqslant \frac{m''}{4\pi^2}\int_{-\pi}^{\pi}|q^*(-um'')|^{-4}\,du\int |q^*(x)|^2\,dx \leqslant \frac{\|q\|^2}{2\pi}\int_{-\pi m''}^{\pi m''}|q^*(-u)|^{-4}\,du.
\end{aligned}$$

We introduce the following notation:

$$\Delta_2(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m}|q^*(u)|^{-4}\,du. \tag{21}$$

Finally, coming back to (20), $\mathrm{Var}_X[V_T(Y_k, Y_{k+1})] \leqslant \|T\|^2\|q\|^2\Delta_2(m'')$ which yields $v \geqslant \|q\|^2\Delta_2(m'')$. Finally, we write $v = \min(\|q\|^2\Delta_2(m''), \Delta^2(m''))$.

We can now use Talagrand's inequality (see Lemma 5):

$$\begin{aligned}
&\mathbb{E}\left[\sup_{T\in B(m,m')}(Z_{n,1}^o)^2(T) - 2(1+2\varepsilon)\frac{2\Delta^2(m'')}{n}\right]_+ \\
&\quad\leqslant \frac{C}{n}\left\{ v e^{-K_1\varepsilon\Delta^2(m'')/v} + \frac{\Delta^2(m'')}{nC^2(\varepsilon)}e^{-K_2C(\varepsilon)\sqrt{\varepsilon}\sqrt{n}}\right\}.
\end{aligned}$$

And then, if $P_1(m,m') \geqslant 4(1+2\varepsilon)\Delta^2(m'')/n$,

$$\sum_{m'\in\mathbb{M}_n}\mathbb{E}\left[\sup_{T\in B(m,m')}(Z_{n,1}^o)^2(T) - P_1(m,m')\right]_+ \leqslant \frac{K}{n}\{I(m) + II(m)\},$$

with $I(m) = \sum_{m'\in\mathbb{M}_n} v e^{-K_1\varepsilon\Delta^2(m'')/v}$; $II(m) = \sum_{m'\in\mathbb{M}_n}(\Delta^2(m'')/nC^2(\varepsilon))e^{-K_2C(\varepsilon)\sqrt{\varepsilon}\sqrt{n}}$.

To bound these terms, we use Lemma 4 which yields to

$$v \leqslant c_3(\pi m'')^{4\gamma+\min(1-s,2-2s)}e^{4b(\pi m'')^s} \quad \text{and} \quad \frac{\Delta^2(m'')}{v} \geqslant c_4(\pi m'')^{(1-s)_+},$$

where c_3 and c_4 depend only on k_0, k_1, γ and s. Therefore,

$$\begin{aligned}
I(m) &\leqslant c_3\sum_{m'\in\mathbb{M}_n}(\pi m'')^{4\gamma+\min(1-s,2-2s)}e^{4b(\pi m'')^s - K_1c_4\varepsilon(\pi m'')^{(1-s)_+}} \\
&\leqslant c_3\sum_{m'\in\mathbb{M}_n}[(\pi m)^{4\gamma+\min(1-s,2-2s)}e^{4b(\pi m)^s} \\
&\quad +(\pi m')^{4\gamma+\min(1-s,2-2s)}e^{4b(\pi m')^s}]e^{-\frac{K_1c_4\varepsilon}{2}[(\pi m)^{(1-s)_+}+(\pi m')^{(1-s)_+}]} \\
&\leqslant c_3(\pi m)^{4\gamma+\min(1-s,2-2s)}e^{4b(\pi m)^s - \frac{K_1c_4\varepsilon}{2}(\pi m)^{(1-s)_+}}\sum_{m'\in\mathbb{M}_n}e^{-\frac{K_1c_4\varepsilon}{2}(\pi m')^{(1-s)_+}} \\
&\quad +c_3e^{-\frac{K_1c_4\varepsilon}{2}(\pi m)^{(1-s)_+}}\sum_{m'\in\mathbb{M}_n}(\pi m')^{4\gamma+\min(1-s,2-2s)}e^{4b(\pi m')^s - \frac{K_1c_4\varepsilon}{2}(\pi m')^{(1-s)_+}}.
\end{aligned}$$

We have to distinguish three cases:

Case $s < (1-s)_+ \Leftrightarrow s < 1/2$: In this case we choose $\varepsilon = 8b/(K_1c_4)$ and then

$$I(m) \leqslant c_3(\pi m)^{4\gamma+1-s} e^{4b[(\pi m)^s-(\pi m)^{(1-s)}]} \sum_{m'\in\mathbb{M}_n} e^{-K_1c_4(\pi m')^{(1-s)}} + c_3 e^{-K_1c_4(\pi m)^{(1-s)}} \sum_{m'\in\mathbb{M}_n} (\pi m')^{4\gamma+1-s} e^{4b[(\pi m')^s-(\pi m')^{(1-s)}]}$$

which implies that $I(m)$ is bounded. Moreover, the definition of $\mathbb{M}_n$ and Lemma 4 give $|\mathbb{M}_n| \leqslant Cn^{\zeta}$ with $C > 0$ and $\zeta > 0$. So $II(m) \leqslant C|\mathbb{M}_n|e^{-K_2'\sqrt{n}}$ is bounded too.

Case $s = (1-s)_+ \Leftrightarrow s = 1/2$: In this case

$$I(m) \leqslant c_3(\pi m)^{4\gamma+1/2} e^{(4b-\frac{K_1c_4\varepsilon}{2})(\pi m)^{1/2}} \sum_{m'\in\mathbb{M}_n} e^{-\frac{K_1c_4\varepsilon}{2}(\pi m')^{1/2}} + c_3 e^{-\frac{K_1c_4\varepsilon}{2}(\pi m)^{1/2}} \sum_{m'\in\mathbb{M}_n} (\pi m')^{4\gamma+1/2} e^{(4b-\frac{K_1c_4\varepsilon}{2})(\pi m')^{1/2}}.$$

We choose ε such that $4b - K_1c_4\varepsilon/2 = -4b$ so that

$$I(m) \leqslant c_3(\pi m)^{4\gamma+1/2} e^{-4b(\pi m)^{1/2}} \sum_{m'\in\mathbb{M}_n} e^{-\frac{K_1c_4\varepsilon}{2}(\pi m')^{1/2}} + c_3 e^{-\frac{K_1c_4\varepsilon}{2}(\pi m)^{1/2}} \sum_{m'\in\mathbb{M}_n} (\pi m')^{4\gamma+1/2} e^{-4b(\pi m')^{1/2}} \leqslant C.$$

The term $II(m)$ is also bounded since ε is a constant.

Case $s > (1-s)_+ \Leftrightarrow s > 1/2$: Here we choose ε such that

$$4b(\pi m'')^s - K_1c_4\varepsilon(\pi m'')^{(1-s)_+}/2 = -4b(\pi m'')^s$$

so that

$$I(m) \leqslant c_3(\pi m)^{4\gamma+\min(1-s,2-2s)} e^{-4b(\pi m)^s} \sum_{m'\in\mathbb{M}_n} e^{-\frac{K_1c_4\varepsilon}{2}(\pi m')^{(1-s)_+}} + c_3 e^{-\frac{K_1c_4\varepsilon}{2}(\pi m)^{(1-s)_+}} \sum_{m'\in\mathbb{M}_n} (\pi m')^{4\gamma+\min(1-s,2-2s)} e^{-4b(\pi m')^s} \leqslant C.$$

Moreover,

$$II(m) \leqslant \sum_{m'\in\mathbb{M}_n} \frac{1}{C^2(\varepsilon)} e^{-K_2\sqrt{8b/K_1c_4}(\pi m'')^{[s-(1-s)_+]/2}\sqrt{n}} \leqslant C.$$

In any case $\varepsilon = [8b/K_1c_4](\pi m'')^{[s-(1-s)_+]_+}$, so that

$$P_1(m, m') = C(q)(\pi m'')^{[s-(1-s)_+]_+}\Delta^2(m'')/n,$$

where $C(q)$ is a constant depending only on k_0, k_1, b, γ, s.

6.5. Proof of Proposition 2

We split $Z_{n,2}(T)$ into two terms:

$$Z_{n,2}(T) = \tfrac{1}{2} Z^o_{n,2}(T) + \tfrac{1}{2} Z^e_{n,2}(T),$$

with

$$\begin{cases} Z^o_{n,2}(T) = \dfrac{2}{n} \displaystyle\sum_{i=1, i \text{ odd}}^{n} \left\{ T(X_i, X_{i+1}) - \int T(x,y) F(x,y)\, dx\, dy \right\}, \\ Z^e_{n,2}(T) = \dfrac{2}{n} \displaystyle\sum_{i=1, i \text{ even}}^{n} \left\{ T(X_i, X_{i+1}) - \int T(x,y) F(x,y)\, dx\, dy \right\}. \end{cases}$$

We bound $\mathbb{E}\left(\left[\sup_{T \in B(m,m')} (Z^o_{n,2})^2(T) - P_2(m,m')\right]_+ \mathbb{1}_{\Omega^*}\right)$. The second term can be bounded in the same way. We write $Z^o_{n,2}(T) = (2/n) \sum_{i=1}^{n/2} \{T(V_i) - \mathbb{E}[T(V_i)]\}$ with $V_i = (X_{2i-1}, X_{2i})$. In order to use Lemma 5, we introduce

$$Z^{o*}_{n,2}(T) = \tfrac{1}{2} v_{n,1}(T) + \tfrac{1}{2} v_{n,2}(T),$$

where

$$\begin{cases} v_{n,1}(T) = \dfrac{1}{p_n} \displaystyle\sum_{l=0}^{p_n - 1} \tfrac{1}{q_n} \sum_{i=2lq_n+1}^{(2l+1)q_n} \left\{ T(V_i^*) - \mathbb{E}[T(V_i^*)] \right\}, \\ v_{n,2}(T) = \dfrac{1}{p_n} \displaystyle\sum_{l=0}^{p_n - 1} \tfrac{1}{q_n} \sum_{i=(2l+1)q_n+1}^{(2l+2)q_n} \left\{ T(V_i^*) - \mathbb{E}[T(V_i^*)] \right\}. \end{cases}$$

Since $X_i = X_i^*$ on Ω^*, we can replace $Z^o_{n,2}$ by $Z^{o*}_{n,2}$. This leads us to bound $\mathbb{E}\left(\left[\sup_{T \in B(m,m')} v^2_{n,1}(T) - P_2(m,m')\right]_+ \mathbb{1}_{\Omega^*}\right)$. So we compute the bounds M_1, H and v of Lemma 5.

(1) If T belongs to $\mathbb{S}_{m''}$, $|T(x,y)|^2 \leqslant \sum_{j,k} a^2_{j,k} \sum_{j,k} \varphi^2_{m'',j}(x) \varphi^2_{m'',k}(y)$ and so

$$\|T\|_\infty \leqslant \|T\| \left\| \sum_j \varphi^2_{m'',j} \right\|_\infty \leqslant \|T\| m'',$$

using (1) of Lemma 3. Then $\|1/q_n \sum_{i=2lq_n+1}^{(2l+1)q_n} T\|_\infty \leqslant \|T\| m''$ and $M_1 = m''$.

(2) Let us compute H^2

$$\sup_{T \in B(m,m')} v^2_{n,1}(T) \leqslant \sum_{j,k} v^2_{n,1}(\varphi_{m'',j} \otimes \varphi_{m'',k}).$$

Then, by taking the expectation,

$$\begin{aligned} \mathbb{E}\left(\sup_{T \in B(m,m')} v^2_{n,1}(T)\right) &\leqslant \sum_{j,k} \frac{1}{p_n^2} \mathrm{Var}\left(\sum_{l=0}^{p_n-1} \frac{1}{q_n} \sum_{i=2lq_n+1}^{(2l+1)q_n} \varphi_{m'',j} \otimes \varphi_{m'',k}(V_i^*) \right) \\ &\leqslant \sum_{j,k} \frac{1}{p_n^2} \sum_{l=0}^{p_n-1} \mathrm{Var}\left(\frac{1}{q_n} \sum_{i=2lq_n+1}^{(2l+1)q_n} \varphi_{m'',j} \otimes \varphi_{m'',k}(V_i^*) \right), \end{aligned}$$

by using independence of the A_l^*. Lemma 6 then gives

$$\mathbb{E}\left(\sup_{T\in B(m,m')} v_{n,1}^2(T)\right) \leqslant \frac{4}{p_n q_n}\left\|\sum_{j,k}(\varphi_{m'',j}\otimes\varphi_{m'',k})^2\right\|_\infty \sum\beta_{2k} \leqslant \frac{16}{n}\left(\sum\beta_{2k}\right)m''^2.$$

Finally, $H = 4\sqrt{\sum\beta_{2k}}m''/\sqrt{n}$.

(3) v remains to be calculated. If T belongs to $B(m')$, using Lemma 6

$$\begin{aligned}\operatorname{Var}\left[\frac{1}{q_n}\sum_{i=2lq_n+1}^{(2l+1)q_n} T(V_i^*)\right] &\leqslant \frac{4}{q_n}\mathbb{E}[T^2(V_1)b(V_1)]\\ &\leqslant \frac{4}{q_n}\|T\|_\infty\sqrt{\mathbb{E}[T^2(V_1)]}\sqrt{\mathbb{E}[b^2(V_1)]}\\ &\leqslant \frac{4}{q_n}\|T\|_\infty\sqrt{\|F\|_\infty}\sqrt{2\sum(k+1)\beta_{2k}}\end{aligned}$$

and so $v = 4\|F\|_\infty^{1/2}\sqrt{2\sum(k+1)\beta_{2k}}m''/q_n$.

By writing Talagrand's inequality (Lemma 5) with $\varepsilon = 1$, we obtain

$$\begin{aligned}&\mathbb{E}\left(\left[\sup_{T\in B(m,m')}(v_{n,1})^2(T) - 6\frac{16}{n}\left(\sum\beta_{2k}\right)m''^2\right]_+ \mathbb{1}_{\Omega^*}\right)\\ &\quad\leqslant \frac{K}{n}\left\{m''e^{-K_1 m''} + \frac{m''q_n^2}{n}e^{-K_2\sqrt{n}/q_n}\right\}.\end{aligned}$$

Then by summation over m'

$$\begin{aligned}&\sum_{m'\in\mathbb{M}_n}\mathbb{E}\left(\left[\sup_{T\in B(m,m')}(v_{n,1})^2(T) - \frac{96}{n}\left(\sum\beta_{2k}\right)m''^2\right]_+ \mathbb{1}_{\Omega^*}\right)\\ &\quad\leqslant \frac{K}{n}\left\{\sum_{m'\in\mathbb{M}_n} m''e^{-K_1 m''} + \sum_{m'\in\mathbb{M}_n} m''n^{2c-1}e^{-K_2 n^{1/2-c}}\right\} \leqslant \frac{C}{n}\end{aligned}$$

since $c < 1/2$. In the same way, we obtain

$$\sum_{m'\in\mathbb{M}_n}\mathbb{E}\left(\left[\sup_{T\in B(m,m')}(v_{n,2})^2(T) - \frac{96}{n}\left(\sum\beta_{2k}\right)m''^2\right]_+ \mathbb{1}_{\Omega^*}\right) \leqslant \frac{C}{n},$$

which yields

$$\sum_{m'\in\mathbb{M}_n}\mathbb{E}\left(\left[\sup_{T\in B(m,m')}(Z_{n,2}^o)^2(T) - P_2(m,m')\right]_+ \mathbb{1}_{\Omega^*}\right) \leqslant \frac{C}{n},$$

with $P_2(m,m') = 96(\sum\beta_{2k})m''^2/n$.

6.6. Proof of Corollary 2

Let us compute the bias term. Since $F_m^* = F^* \mathbb{1}_{[-\pi m, \pi m]^2}$,

$$\|F - F_m\|^2 = \frac{1}{4\pi^2} \iint_{([-\pi m,\pi m]^2)^c} |F^*(u,v)|^2 \, du \, dv$$
$$\leqslant \frac{1}{4\pi^2} \iint_{[-\pi m,\pi m]^c \times \mathbb{R}} |F^*(u,v)|^2 \, du \, dv + \frac{1}{4\pi^2} \iint_{\mathbb{R} \times [-\pi m,\pi m]^c} |F^*(u,v)|^2 \, du \, dv.$$

But

$$\iint_{[-\pi m,\pi m]^c \times \mathbb{R}} |F^*(u,v)|^2 \, du \, dv \leqslant L((\pi m)^2 + 1)^{-\Delta} e^{-2A(\pi m)^R}.$$

Thus $\|F - F_m\|^2 = O((\pi m)^{-2\Delta} e^{-2A(\pi m)^R})$ and

$$\mathbb{E}\|F - \tilde{F}\|^2 \leqslant C' \inf_{m \in \mathbb{M}_n} \left\{ (\pi m)^{-2\Delta} e^{-2A(\pi m)^R} + (\pi m)^{[s-(1-s)_+]_+ + 4\gamma + 2 - 2s} \frac{e^{4b(\pi m)^s}}{n} \right\} + \frac{C}{n}.$$

Next the bias-variance trade-off is performed similarly to Lacour [22].

6.7. Proof of Theorem 3

Let $E_n = \{\|f - \tilde{f}\|_\infty \leqslant f_0/2\}$. On E_n and for $x \in B$, $\tilde{f}(x) = \tilde{f}(x) - f(x) + f(x) \geqslant f_0/2$. Since $\tilde{F}$ belongs to $\mathbb{S}_{\hat{M}}$, using (2), $\|\tilde{F}\|_\infty \leqslant \hat{M}\|\tilde{F}\|$. Now (18) gives $\|\tilde{F}\| \leqslant \Delta(\hat{M})$ so that $\|\tilde{F}\|_\infty \leqslant \hat{M}\Delta(\hat{M})$. Since $\hat{M}$ belongs to $\mathbb{M}_n$, $\Delta(\hat{M}) \leqslant \sqrt{n}$ and Lemma 4 gives $\hat{M} \leqslant C\Delta(\hat{M})^{1/(2\gamma+1)}$ if $s = 0$ or $\hat{M} \leqslant C(\ln \Delta(\hat{M}))^{1/s}$ otherwise. So, for n large enough, $(2/f_0)\|\tilde{F}\|_\infty \leqslant n$ and $\tilde{\Pi}(x,y) = \tilde{F}(x,y)/\tilde{f}(x)$.

For all $(x,y) \in B^2$,

$$|\tilde{\Pi}(x,y) - \Pi(x,y)|^2 \leqslant \left| \frac{\tilde{F}(x,y) - \tilde{f}(x)\Pi(x,y)}{\tilde{f}(x)} \right|^2 \mathbb{1}_{E_n} + (|\tilde{\Pi}(x,y)| + |\Pi(x,y|)^2 \mathbb{1}_{E_n^C}$$
$$\leqslant \frac{|\tilde{F}(x,y) - F(x,y) + \Pi(x,y)(f(x) - \tilde{f}(x))|^2}{f_0^2/4}$$
$$+ 2(\|\tilde{\Pi}\|_\infty^2 + |\Pi(x,y)|^2) 1_{E_n^C}.$$

Since $\int_B \Pi^2(x,y)\,dy \leqslant \|\Pi\|_{B,\infty} \int_B \Pi(x,y)\,dy \leqslant \|\Pi\|_{B,\infty}$ for all $x \in B$,

$$\mathbb{E}\|\Pi - \tilde{\Pi}\|_B^2 \leqslant \frac{8}{f_0^2}[\mathbb{E}\|F - \tilde{F}\|^2 + \|\Pi\|_{B,\infty}\mathbb{E}\|f - \tilde{f}\|^2] + 2|B|(|B|n^2 + \|\Pi\|_{B,\infty})P(E_n^C).$$

We still have to prove that $P(E_n^C) \leqslant Cn^{-3}$. Given that $\|f - \tilde{f}\|_\infty \leqslant \|f - f_{\hat{m}}\|_\infty + \|f_{\hat{m}} - \hat{f}_{\hat{m}}\|_\infty$ we obtain

$$P(E_n^c) \leqslant P(\|f - f_{\hat{m}}\|_\infty > f_0/4) + P(\|f_{\hat{m}} - \hat{f}_{\hat{m}}\|_\infty > f_0/4).$$

Let us prove now that if f belongs to $\mathcal{A}_{\delta,r,a}(l)$, $\|f - f_m\|_\infty = O(m^{1/2-\delta-r/2} e^{-a(\pi m)^r})$. Since $f_m^* = f^* \mathbb{1}_{[-\pi m,\pi m]}$ and using the inverse Fourier transform,

$$\|f - f_m\|_\infty \leqslant \frac{1}{2\pi} \int_{|u| \geqslant \pi m} |f^*(u)|\,du.$$

If $r > 0$, let $0 < \alpha < a$. By considering that function $x \mapsto (x^2+1)^{\delta/2} e^{(a-\alpha)|x|^r}$ is increasing and using the Schwarz inequality, we obtain

$$\|f - f_m\|_\infty \leqslant \frac{1}{2\pi}((\pi m)^2+1)^{-\delta/2} e^{(\alpha-a)(\pi m)^r} \sqrt{l} \sqrt{\int_{|u|\geqslant \pi m} e^{-2\alpha|u|^r}\, du}.$$

But $\int_{|u|\geqslant \pi m} e^{-2\alpha|u|^r} \leqslant C(\pi m)^{1-r} e^{-2\alpha(\pi m)^r}$ and then

$$\begin{aligned}\|f - f_m\|_\infty &\leqslant \frac{\sqrt{Cl}}{2\pi}((\pi m)^2+1)^{-\delta/2} e^{(\alpha-a)(\pi m)^r} (\pi m)^{1/2-r/2} e^{-\alpha(\pi m)^r}\\ &= O(m^{1/2-\delta-r/2} e^{-a(\pi m)^r})\end{aligned}$$

and this is still valid when $r = 0$. Thus, since $\hat{m} \geqslant \ln\ln n$, $\|f - f_{\hat{m}}\|_\infty \to 0$ and for n large enough $P(\|f - f_{\hat{m}}\|_\infty > f_0/4) = 0$. Next

$$P(\|f_{\hat{m}} - \hat{f}_{\hat{m}}\|_\infty > f_0/4) \leqslant P(\Omega^{*c}) + P\left(\|f_{\hat{m}} - \hat{f}_{\hat{m}}\| \mathbb{1}_{\Omega^*} > \frac{f_0}{4\sqrt{\hat{m}}}\right).$$

Since $c\theta > 4$, $P(\Omega^{*c}) \leqslant Mn^{1-c\theta} \leqslant Mn^{-3}$. We still have to prove that

$$P\left(\|f_{\hat{m}} - \hat{f}_{\hat{m}}\| \mathbb{1}_{\Omega^*} > \frac{f_0}{4\sqrt{\hat{m}}}\right) \leqslant \frac{C}{n^3}.$$

First, we observe that

$$\|f_{\hat{m}} - \hat{f}_{\hat{m}}\|^2 = \sum_{j\in\mathbb{Z}} \left(\frac{1}{n}\sum_{i=1}^n v_{\varphi_{\hat{m}j}}(Y_i) - \mathbb{E}[v_{\varphi_{\hat{m}j}}(Y_i)]\right)^2 = \sup_{t\in B_{\hat{m}}} \nu_n^2(t)$$

where $\nu_n(t) = \frac{1}{n}\sum_{i=1}^n v_t(Y_i) - \mathbb{E}[v_t(Y_i)]$, $B_m = \{t \in S_m, \|t\| \leqslant 1\}$.

Then

$$P\left(\|f_{\hat{m}} - \hat{f}_{\hat{m}}\| \mathbb{1}_{\Omega^*} > \frac{f_0}{4\sqrt{\hat{m}}}\right) = P\left(\sup_{t\in B_{\hat{m}}} |\nu_n(t)| \mathbb{1}_{\Omega^*} > \frac{f_0}{4\sqrt{\hat{m}}}\right).$$

As previously, we split $\nu_n(t)$ into two terms

$$\begin{aligned}\nu_n(t) = &\frac{1}{2p_n}\sum_{l=0}^{p_n-1} \frac{1}{q_n}\sum_{i=2lq_n+1}^{(2l+1)q_n} v_t(Y_i) - \mathbb{E}[v_t(Y_i)]\\ &+\frac{1}{2p_n}\sum_{l=0}^{p_n-1} \frac{1}{q_n}\sum_{i=(2l+1)q_n+1}^{(2l+2)q_n} v_t(Y_i) - \mathbb{E}[v_t(Y_i)]\end{aligned}$$

and it is sufficient to study

$$P\left(\sup_{t\in B_{\hat{m}}} \left|\frac{1}{p_n}\sum_{l=0}^{p_n} \frac{1}{q_n}\sum_{i=2lq_n+1}^{(2l+1)q_n} v_t(Y_i^*) - \mathbb{E}[v_t(Y_i^*)]\right| > \frac{f_0}{4\sqrt{\hat{m}}}\right).$$

We bound this term by the sum

$$\sum_{m\in\mathcal{M}_n} P\left(\sup_{t\in B_m} \left|\frac{1}{p_n}\sum_{l=0}^{p_n} \frac{1}{q_n}\sum_{i=2lq_n+1}^{(2l+1)q_n} v_t(Y_i^*) - \mathbb{E}[v_t(Y_i^*)]\right| > \frac{f_0}{4\sqrt{m}}\right)$$

and we use inequality (22) in proof of Lemma 5 with $\eta = 1$ and $\lambda = \frac{f_0}{8\sqrt{m}}$:

$$P\left(\sup_{t\in B_m}\left|\frac{1}{p_n}\sum_{l=0}^{p_n}\frac{1}{q_n}\sum_{i=2lq_n+1}^{(2l+1)q_n} v_t(Y_i^*) - \mathbb{E}[v_t(Y_i^*)]\right| > 2H + \lambda\right)$$
$$\leqslant \exp\left(-K p_n \min\left(\frac{\lambda^2}{v}, \frac{\lambda}{M_1}\right)\right).$$

Here, we compute

$$M_1 = \sqrt{\Delta(m)}, \quad H^2 = 8\sum_k \beta_k \frac{\Delta(m)}{n}, \quad v = 4\sum_k \beta_k \frac{\Delta(m)}{q_n}.$$

Thus

$$P\left(\sup_{t\in B_m}|v_n(t)| > 2H + \frac{f_0}{8\sqrt{m}}\right) \leqslant 2\exp\left(-K' \min\left(\frac{n}{m\Delta(m)}, \frac{p_n}{\sqrt{m\Delta(m)}}\right)\right).$$

Now we use the assumption $\forall m\; m\Delta(m) \leqslant n/(\ln n)^2$. For n large enough, $2H = 4\sqrt{2\sum_k \beta_k}\sqrt{\Delta(m)}/\sqrt{n} \leqslant f_0/(8\sqrt{m})$. So

$$\sum_{m\in\mathcal{M}_n} P\left(\sup_{t\in B_m}|v_n(t)| > \frac{f_0}{4\sqrt{m}}\right) \leqslant 2|\mathcal{M}_n| \exp\left(-K' \min\left((\ln n)^2, n^{1/2-c}\ln n\right)\right) \leqslant \frac{C}{n^3}.$$

6.8. Technical lemmas

Lemma 3. *For each $m \in \mathcal{M}_n$*

(1) $\|\sum_j \varphi_{m,j}^2\|_\infty = m$,
(2) $v_{\varphi_{m,j}}(x) = \sqrt{m}/(2\pi)\int_{-\pi}^{\pi} e^{-ijv}e^{ixvm}[q^*(-vm)]^{-1}\,dv$,
(3) $\|\sum_j |v_{\varphi_{m,j}}|^2\|_\infty = \Delta(m)$,

where $\Delta(m)$ is defined in (6).

Proof of Lemma 3. First we remark that

$$\varphi_{m,j}^*(u) = \int e^{-ixu}\sqrt{m}\varphi(mx - j)\,dx$$
$$= \frac{1}{\sqrt{m}}e^{-iju/m}\int e^{-ixu/m}\varphi(x)\,dx = \frac{1}{\sqrt{m}}e^{-iju/m}\varphi^*\left(\frac{u}{m}\right).$$

Thus, using the inverse Fourier transform

$$\varphi_{m,j}(x) = \frac{1}{2\pi}\int e^{iux}\frac{1}{\sqrt{m}}e^{-iju/m}\varphi^*\left(\frac{u}{m}\right)du = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ijv}\sqrt{m}e^{ixvm}\,dv.$$

The Parseval equality yields $\sum_j \varphi_{m,j}^2(x) = 1/2\pi \int_{-\pi}^{\pi} |\sqrt{m} e^{ixvm}|^2\, dv = m$. The first point is proved. Now we compute $v_{\varphi_{m,j}}(x)$

$$v_{\varphi_{m,j}}(x) = \frac{1}{2\pi}\int e^{ixu}\frac{\varphi_{m,j}^*(u)}{q^*(-u)}\, du = \frac{1}{2\pi}\int e^{ixu}\frac{1}{\sqrt{m}}e^{-iju/m}\varphi^*\left(\frac{u}{m}\right)\frac{du}{q^*(-u)}$$
$$= \frac{\sqrt{m}}{2\pi}\int e^{-ijv}e^{ixvm}\frac{\varphi^*(v)}{q^*(-vm)}\, dv.$$

But $\varphi^*(v) = \mathbb{1}_{[-\pi,\pi]}(v)$ and thus the second point is proved. Moreover, $v_{\varphi_{m,j}}(x)$ can be seen as a Fourier coefficient. Parseval's formula then gives

$$\sum_j |v_{\varphi_{m,j}}(x)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sqrt{m}e^{ixvm}\frac{1}{q^*(-vm)}\right|^2 dv = \frac{m}{2\pi}\int \left|q^*(-vm)\right|^{-2} dv.$$

Therefore, $\| \sum_j |v_{\varphi_{m,j}}|^2 \|_\infty = 1/2\pi \int_{-\pi m}^{\pi m} |q^*(-u)|^{-2}\, du = \Delta(m)$. □

Lemma 4. *If q verifies $|q^*(x)| \geqslant k_0(x^2+1)^{-\gamma/2}\exp(-b|x|^s)$, then*

(1) $\Delta(m) \leqslant c_1(\pi m)^{2\gamma+1-s}e^{2b(\pi m)^s}$,
(2) $\Delta_2(m) \leqslant c_2(\pi m)^{4\gamma+1-s}e^{4b(\pi m)^s}$.

Moreover if $|q^*(x)| \leqslant k_1(x^2+1)^{-\gamma/2}\exp(-b|x|^s)$, then $\Delta(m) \geqslant c_1'(\pi m)^{2\gamma+1-s}e^{2b(\pi m)^s}$.

The proof of this result is omitted. It is obtained by distinguishing the cases $s > 2\gamma+1$ and $s \leqslant 2\gamma+1$ and with standard evaluations of integrals.

Lemma 5. *Let $T_1, \ldots, T_n$ be independent random variables and $\nu_n(r) = (1/n)\sum_{i=1}^n [r(T_i) - \mathbb{E}(r(T_i))]$, for r belonging to a countable class $\mathcal{R}$ of measurable functions. Then, for $\varepsilon > 0$,*

$$\mathbb{E}\left[\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - 2(1+2\varepsilon)H^2\right]_+ \leqslant C\left(\frac{v}{n}e^{-K_1\varepsilon\frac{nH^2}{v}} + \frac{M_1^2}{n^2C^2(\varepsilon)}e^{-K_2C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{M_1}}\right)$$

with $K_1 = 1/6$, $K_2 = 1/(21\sqrt{2})$, $C(\varepsilon) = \sqrt{1+\varepsilon}-1$ and C a universal constant and where

$$\sup_{r\in\mathcal{R}} \|r\|_\infty \leqslant M_1, \quad \mathbb{E}\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|\right) \leqslant H, \quad \sup_{r\in\mathcal{R}} \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(r(T_i)) \leqslant v.$$

Usual density arguments allow to use this result with noncountable class of functions $\mathcal{R}$.

Proof of Lemma 5. We apply the Talagrand concentration inequality given in Klein and Rio [21] to the functions $s^i(x) = r(x) - \mathbb{E}(r(T_i))$ and we obtain

$$P\left(\sup_{r\in\mathcal{R}} |\nu_n(r)| \geqslant H + \lambda\right) \leqslant \exp\left(-\frac{n\lambda^2}{2(v+4HM_1)+6M_1\lambda}\right).$$

Then we modify this inequality following Birgé and Massart [5, Corollary 2, p. 354] . It gives

$$P\left(\sup_{r\in\mathcal{R}} |\nu_n(r)| \geqslant (1+\eta)H + \lambda\right) \leqslant \exp\left(-\frac{n}{3}\min\left(\frac{\lambda^2}{2v}, \frac{\min(\eta,1)\lambda}{7M_1}\right)\right). \tag{22}$$

To conclude we set $\eta = \sqrt{1+\epsilon} - 1$ and we use the formula $\mathbb{E}[X]_+ = \int_0^\infty P(X \geqslant t)\,dt$ with $X = \sup_{r\in\mathcal{R}} |v_n(r)|^2 - 2(1+2\epsilon)H^2$. $\square$

Lemma 6 (*Viennet [32, Theorem 2.1 and Lemma 4.2]*). *Let (T_i) a strictly stationary process with β-mixing coefficients β_k. Then there exists a function b such that*

$$\mathbb{E}[b(T_1)] \leqslant \sum_k \beta_k \quad \textit{and} \quad \mathbb{E}[b^2(T_1)] \leqslant 2\sum_k (k+1)\beta_k$$

and for all function ψ (such that $\mathbb{E}[\psi^2(T_1)] < \infty$) and for all N

$$\mathrm{Var}\left(\sum_{i=1}^N \psi(T_i)\right) \leqslant 4N\mathbb{E}[\psi^2(T_1)b(T_1)].$$

References

[1] D. Bakry, X. Milhaud, P. Vandekerkhove, Statistique de chaînes de Markov cachées à espace d'états fini. Le cas non stationnaire, C. R. Acad. Sci. Paris Sér. I. Math. 325 (2) (1997) 203–206.
[2] A. Barron, L. Birgé, P. Massart, Risk bounds for model selection via penalization, Probab. Theory Related Fields 113 (3) (1999) 301–413.
[3] L.E. Baum, T. Petrie, Statistical inference for probabilistic functions of finite state Markov chains, Ann. Math. Statist. 37 (1966) 1554–1563.
[4] P.J. Bickel, Y. Ritov, T. Rydén, Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models, Ann. Statist. 26 (4) (1998) 1614–1635.
[5] L. Birgé, P. Massart, Minimum contrast estimators on sieves: exponential bounds and rates of convergence, Bernoulli 4 (3) (1998) 329–375.
[6] C. Butucea, Deconvolution of supersmooth densities with smooth noise, Canad. J. Statist. 32 (2) (2004) 181–192.
[7] C. Butucea, C. Matias, Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model, Bernoulli 11 (2) (2005) 309–340.
[8] C. Butucea, A.B. Tsybakov, Sharp optimality for density deconvolution with dominating bias, Theory Probab. Appl. 52 (1) (2007), to appear.
[9] O. Cappé, E. Moulines, T. Rydén, Inference in hidden Markov models, Springer Series in Statistics, Springer, New York, 2005.
[10] R.J. Carroll, P. Hall, Optimal rates of convergence for deconvolving a density, J. Amer. Statist. Assoc. 83 (404) (1988) 1184–1186.
[11] M. Chaleyat-Maurel, V. Genon-Catalot, Computable infinite dimensional filters with applications to discretized diffusions, Stochastic Process. Appl. 116 (10) (2006) 1447–1467.
[12] S. Clémençon, Nonparametric estimation for some specific classes of hidden Markov models, Preprint Modal'X n° 03-9; 2003, ⟨http://www.u-paris10.fr/65897276/0/fiche___pagelibre/⟩.
[13] F. Comte, J. Dedecker, M.-L. Taupin, Adaptive density deconvolution with dependent input, Preprint MAP5 n° 2006-4; 2006a, ⟨http://www.math-info.univ-paris5.fr/map5/publis/titres06.html⟩.
[14] F. Comte, Y. Rozenholc, M.-L. Taupin, Penalized contrast estimator for adaptive density deconvolution, Canad. J. Statist. 34 (3) (2006) 431–452.
[15] C.C.Y. Dorea, L.C. Zhao, Nonparametric density estimation in hidden Markov models, Statist. Inference Stochastic. Process. 5 (1) (2002) 55–64.
[16] R. Douc, É. Moulines, T. Rydén, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, Ann. Statist. 32 (5) (2004) 2254–2304.
[17] P. Doukhan, Mixing. Properties and Examples, Lecture Notes in Statistics, vol. 85, Springer, New York, 1994.
[18] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist. 19 (3) (1991) 1257–1272.
[19] V. Genon-Catalot, T. Jeantheau, C. Larédo, Stochastic volatility models as hidden Markov models and statistical applications, Bernoulli 6 (6) (2000) 1051–1079.
[20] J.L. Jensen, N.V. Petersen, Asymptotic normality of the maximum likelihood estimator in state space models, Ann. Statist. 27 (2) (1999) 514–535.

[21] T. Klein, E. Rio, Concentration around the mean for maxima of empirical processes, Ann. Probab. 33 (3) (2005) 1060–1077.
[22] C. Lacour, Rates of convergence for nonparametric deconvolution, C. R. Acad. Sci. Paris Sér. I Math. 342 (11) (2006) 877–882.
[23] C. Lacour, Adaptive estimation of the transition density of a Markov chain, Ann. Inst. H. Poincaré Probab. Statist, 2007, available online, to appear.
[24] O.V. Lepski, B.Y. Levit, Adaptive nonparametric estimation of smooth multivariate functions, Math. Methods Statist. 8 (3) (1999) 344–370.
[25] B.G. Leroux, Maximum-likelihood estimation for hidden Markov models, Stochastic Process. Appl. 40 (1) (1992) 127–143.
[26] E. Masry, Strong consistency and rates for deconvolution of multivariate densities of stationary processes, Stochastic Process. Appl. 47 (1) (1993) 53–74.
[27] C. Matias, Semiparametric deconvolution with unknown noise variance, ESAIM Probab. Statist. 6 (2002) 271–292 (electronic), new directions in time series analysis (Luminy, 2001).
[28] M. Pensky, B. Vidakovic, Adaptive wavelet estimator for nonparametric density deconvolution, Ann. Statist. 27 (6) (1999) 2033–2053.
[29] L.A. Stefanski, Rates of convergence of some estimators in a class of deconvolution problems, Statist. Probab. Lett. 9 (3) (1990) 229–235.
[30] K. Tribouley, G. Viennet, L_p adaptive density estimation in a β mixing framework, Ann. Inst. H. Poincaré Probab. Statist. 34 (2) (1998) 179–208.
[31] B. Van Es, P. Spreij, H. van Zanten, Nonparametric volatility density estimation for discrete time models, J. Nonparametric Statist. 17 (2) (2005) 237–251.
[32] G. Viennet, Inequalities for absolutely regular sequences: application to density estimation, Probab. Theory Related Fields 107 (4) (1997) 467–492.